DOCUMENT RESUME

ED 148 122 FL 009 072

AUTHOR Zerling, J. P.
TITLE L'Ordinateur, Un Nouveau Pedagogue? Quelques Exemples d'Application de l'Ordinateur a l'Enseignement des Langues Vivantes (The Computer: A New Teacher? Some Examples of the Use of the Computer in Second Language Teaching). Melanges Pedagogiques, 1975.
INSTITUTION Nancy Univ., (France). Centre de Recherches et d'Applications Pedagogiques en Langues.
PUB DATE 75
NOTE 25p.; In French ; Contains occasional small type

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Computer Assisted Instruction; Computers; Computer Science; Educational Technology; Grades (Scholastic); Instructional Aids; Instructional Materials; *Language Instruction; Language Skills; Language Tests; Programed Instruction; *Reading Comprehension; *Second Language Learning; *Statistical Analysis; Student Evaluation; Teaching Methods; *Test Construction

ABSTRACT

The aim of this paper is to present linguists--both members of research teams and language teachers--with some possible applications of computing science to language teaching. First, it is possible to use a computer's output system merely in order to print any type of information, e.g., a list of students' names and their marks. The computer also allows the teacher to make more elaborate calculations and obtain very useful results, such as the arithmetic mean, the standard deviation and so on, whatever the number of students. To illustrate clearly these applications, a concrete example is given concerning the improvement of a reading comprehension test. The test is fully described and its evaluation over three years is presented and explained: modifications of all sorts were introduced in accordance with the interpretation of mathematical and statistical data. The discrimination factor makes it clear whether any particular item discriminates effectively; the frequency factor shows the number of answers--right or wrong--given for each exercise. Finally, the correlation factor shows whether all the different series of exercises really contribute to testing the same kind of reading skill. All these data are plotted or given in charts, and are thoroughly discussed in order to explain how it is possible to transform such skill-testing material to make it as concise as possible while giving it as great a discriminatory power as possible. (Author/CLK)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

ED148122

MELANGES PEDAGOGIQUES 1975

J.-P. ZERLING

L'ORDINATEUR, UN NOUVEAU PEDAGOGUE ?

Quelques exemples d'application de l'ordinateur à l'enseignement des langues vivantes

ED009072

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

C.R.A.P.E.L.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

C.R.A.P.E.L.

INTRODUCTION

La possibilité d'utiliser depuis peu à Nancy un ordinateur dans le cadre de la recherche en linguistique nous a fait nous pencher sur certaines des applications possibles du calcul automatique à l'enseignement des langues vivantes.

Cette curiosité du chercheur s'est vue stimulée en 1973-74 par la nécessité de posséder un moyen rapide de manipuler une masse importante d'informations : en effet, la mise en place récente du DEUG [1] a entraîné dès 1973 l'apparition sur le « marché » des langues vivantes d'environ 120 étudiants dépendant de l'Université de Nancy II et ayant choisi l'anglais comme langue vivante (élément 5 %). En 1974, leur nombre était d'environ 400 et il sera en 1975 compris entre 500 et 600.

Nous présentons dans cet article trois exemples d'application pratique de l'utilisation de l'ordinateur pour le traitement des données :
— Traitement administratif.
— Exploitation pédagogique.
— Optimisation de matériaux pédagogiques (Description d'un test de compréhension écrite et de sa mise au point).

Notre propos n'est pas ici de présenter un ensemble de renseignements d'ordre informatique, statistique ou mathématique, mais plutôt de citer quelques exemples de traitement automatique actuellement entrepris à Nancy II. Nous renvoyons donc à des ouvrages plus spécialisés le lecteur qui serait intéressé par plus de précisions quant à certains termes utilisés dans cet article.

[1] DEUG : Diplôme d'Enseignement Universitaire Général.

Avant d'entrer dans le détail de ces applications, il paraît nécessaire de décrire très brièvement le matériel pédagogique dont nous disposons :

— en début d'année, les étudiants sont soumis à un **test de classement** créé spécialement pour évaluer leur niveau de compréhension écrite. Ce test est décrit avec plus de précision dans la suite de l'article. L'ensemble des résultats pour chaque étudiant est entièrement codifiable, tient facilement sur une carte d'ordinateur (80 caractères), et peut donc aisément être soumis à un traitement, à une correction ou à une analyse automatique.

— en cours d'année et pour l'examen final, aucune correction automatique n'a encore été envisagée pour l'instant, mais les notes successives obtenues par chaque étudiant sont également perforées sur carte ainsi que ses nom, prénoms et UER.

Avant d'exposer les avantages que peut représenter un ordinateur à un enseignant en possession de telles données, signalons encore au lecteur qui ne serait pas familiarisé avec les problèmes de l'informatique que l'ordinateur n'est un cerveau que dans la mesure où il est capable d'effectuer des calculs d'une manière extrêmement rapide. Pour donner un ordre d'idée, précisons que l'ordinateur que nous utilisons à l'Institut Universitaire de Calcul Automatique (IUCA) de Nancy II (CII 10070) a un temps de calcul pour une opération élémentaire (addition ou soustraction) de l'ordre de la microseconde. Toutes les opérations ou manipulations mentionnées dans la suite de cet article seraient donc réalisables « manuellement », mais elles nécessiteraient un temps de travail qu'aucun chercheur ne pourrait décemment leur consacrer.

1. TRAITEMENT ADMINISTRATIF

Nous entendons par traitement administratif la manipulation pure et simple de l'ensemble des notes des étudiants, sans aucune intention pédagogique. C'est donc ici le facteur *rapidité* de l'ordinateur auquel nous faisons appel. La mise au point de divers programmes relativement simples en langage FORTRAN [2] nous permet d'effectuer à l'heure actuelle les opérations suivantes :

— pour le test de classement : calcul automatique des notes et écriture des résultats classés ou non classés.

— pour l'ensemble des autres notes, nous avons le choix entre les traitements suivants :

- — calcul automatique de la moyenne de plusieurs notes affectées ou non d'un coefficient.
- — classement des étudiants par ordre alphabétique.
- — classement des étudiants en fonction de leur note.
- — classement selon des niveaux fixés arbitrairement (découpage en diverses classes : débutants, avancés, etc.).
- — classement par section ou par UER.

Il est évidemment possible de combiner à souhait ces divers programmes, et l'on peut obtenir, par exemple, une liste décroissante des étudiants en psychologie ayant obtenu une moyenne comprise entre 40 et 60 sur 100.

L'intérêt de l'utilisation de tels programmes est certain : on peut obtenir en quelques instants, et sans erreur, des listes de noms ou de résultats qui demanderaient sans cela des heures à être dactylographiées et vérifiées.

[2] Tous les programmes d'ordinateur dont il est question dans la suite de l'article ont été écrits et utilisés par l'auteur.

2. EXPLOITATION PEDAGOGIQUE

Parallèlement à ce traitement rapide des notes, on peut aussi procéder à une exploitation statistique simple des résultats, qui peut permettre de tirer des conclusions d'ordre pédagogique. C'est ici non seulement le facteur rapidité de l'ordinateur qui est utilisé, mais la possibilité de lui faire effectuer des *opérations plus sophistiquées*, plus longues, telles qu'extraction de racines, élévation à la puissance, etc.

On peut ainsi, pour chaque série de notes, calculer au moins deux valeurs : la **moyenne arithmétique** et l' **écart type** [3]. La connaissance de ces deux nombres donne une assez bonne idée de la répartition des étudiants en fonction de leur note : La moyenne est utile pour éviter de tomber dans le piège très subjectif qui tend à faire comparer toute note à la moyenne *idéale* (par exemple : 10/20 ou 50/100). L'écart type permet pour sa part d'évaluer la dispersion des notes autour de la moyenne : un écart type faible correspond à des notes regroupées autour de la moyenne, alors qu'un écart type élevé indiquera un étalement des notes (cf. figures 3b et 3c).

Parallèlement à cette manipulation des chiffres, l'ordinateur permet également une écriture rapide, et conséquemment une représentation graphique des résultats, par exemple une représentation de la répartition des étudiants en fonction de leur note (figure 1). Ainsi, à partir de tracés, l'enseignant peut plus aisément juger sa façon de noter, le bien fondé des barèmes utilisés, ou la qualité des exercices proposés. L'utilité de la représentation graphique se fait tout particulièrement sentir quand on détermine une moyenne à partir de plusieurs autres notes : c'est le cas lorsqu'on calcule une moyenne de fin d'année, ou bien la note globale d'une batterie de tests. Un exemple simple peut démontrer l'utilité d'un graphique : les tracés A et B de la figure 2 représentent les répartitions d'un groupe d'étudiants à deux épreuves, la première étant apparemment plus difficile que la seconde. Si l'on fait pour chaque étudiant la moyenne arithmétique des deux notes, un étudiant E1 ayant échoué à l'épreuve A et moyennement réussi à l'épreuve B n'aura pas la même note finale qu'un étudiant E2 ayant travaillé de façon inverse (moyen à A, échec à B). Sera désavantagé l'étudiant E2 qui aura le mieux réussi l'épreuve la plus difficile, situation pour le moins aberrante ! Une représentation graphique peut donc aider à prendre conscience de tels problèmes.

[3] Pour éviter toute ambiguïté, nous donnons ici une formule de calcul pour chacune de ces valeurs :

$$m = \frac{\sum x}{A} \qquad s = \sqrt{\frac{\sum (x-m)^2}{N}}$$

où : m = moyenne arithmétique, x = valeur d'une note de la série, s = écart type, N = nombre de notes dans la série.

3. EXEMPLE D'OPTIMISATION D'UN TEST DE CLASSEMENT

Nous entendons par test de classement en langue vivante un ensemble d'exercices dont la note globale doit permettre de classer des individus en fonction de leur aptitude à l'apprentissage d'un aspect ou de plusieurs aspects de la langue. Par exemple : compréhension écrite, expression écrite, compréhension orale, etc.

Le problème que nous posons est le suivant : lorsqu'un test a été créé, quels sont les critères qui permettent de décider de sa validité ? Nous ne voyons qu'une seule manière d'apporter une réponse à cette question : c'est l'exploitation systématique des résultats obtenus par les personnes ayant subi le test. Or, de tous les enseignants qui créent ou utilisent ce genre de matériel pédagogique, seul un petit nombre, nous le craignons, a la possibilité d'analyser les résultats, et ceci pour de multiples raisons : parfois le nombre d'étudiants est statistiquement insuffisant, ou au contraire trop élevé pour permettre un traitement manuel. Parfois, l'enseignant a tout simplement la conviction que son test correspond entièrement à son attente et que les résultats sont valables. Enfin, dans d'autres cas, c'est simplement un manque d'information qui arrête l'enseignant. Comment peut-il vérifier d'une manière *objective* qu'un test fonctionne bien dans le sens désiré ? Car à notre époque, la simple constatation *subjective* de l'efficacité d'un outil de travail paraît désormais insuffisante.

Certes, quelques travaux de recherche ont déjà été entrepris à Nancy II pour mesurer l'efficacité de certains tests couramment utilisés (cf., par exemple, Lonchamp [1971]), mais nous pensons que ce type de recherche est actuellement insuffisant et qu'il devrait revêtir un caractère beaucoup plus systématique.

Le test de classement que nous avons créé à Nancy II[4] dans l'optique de l'enseignement de l'anglais aux étudiants non-spécialistes du DEUG n'est certainement pas le premier du genre car les tests sont désormais des instruments de travail fréquemment utilisés, mais il existe pourtant plusieurs différences entre celui-ci et la plupart des autres. Tout d'abord, il a été créé consciemment dans l'optique d'une correction et d'un traitement automatiques, puis il a été modifié au fil du temps en fonction de l'analyse statistique de ses résultats : une première version a été soumise en 1973 à environ 120 étudiants ; une deuxième version a été soumise en 1974 à environ 400 étudiants ; et une troisième version, peut-être définitive sera prête pour 1975.

Voyons maintenant en quoi consiste ce test, et de quelle manière nous l'avons peu à peu modifié.

[4] Ont participé à l'élaboration de ce test : Mesdames D. ABE, J. BILLANT, H. ILARI-FINCH, J. THARIAT et Monsieur J.-P. ZERLING.

3.1. Présentation du test

Dans sa première version, ce test reprenait quelques éléments d'un test déjà utilisé par le CRAPEL [5], réalisé par M. Michel KUHN qui avait bien voulu nous laisser emprunter quelques-uns de ses exercices. Le reste des exercices a été créé de toutes pièces par notre équipe de travail comportant, il va sans dire, une personne de langue anglaise.

Ce test est en fait constitué d'une batterie de 7 séries d'exercices différents qui ont pour but unique d'évaluer le **niveau de compréhension écrite** en rejetant dans la mesure du possible toute notion d'expression écrite ou de connaissances orales.

Les séries A, B et C consistent à choisir un élément parmi 4 ou 5 autres éléments fournis (réponses à choix multiple).

A) Choix de la traduction exacte d'un mot composé.

Exemple : an apple pie

a) une cravate couleur pomme
b) une pomme d'api
c) une tarte aux pommes (x)
d) une pie voleuse

B) Choix d'un élément sémantiquement ou grammaticalement identique à un autre.

Exemple : The prisoner stood with a policeman on **either** side.

a) or
b) other
c) every
d) both
e) each (x)

C) Choix d'un élément sémantiquement différent des autres :

Exemple :

a) far more
b) any more (x)
c) much more
d) considerably more

Dans les séries D et E, il ne s'agit plus de choisir un mot dans une liste, mais dans une phrase :

D) Trouver le mot auquel se rapporte un élément souligné dans une phrase.

Exemple : People work harder. They **have to** if they want to survive.

(Réponse : work harder)

[5] Test de compréhension écrite pour non-spécialistes (GRAP, 1968).

E) Déterminer le verbe ou l'absence de verbe dans des titres de journaux.
Exemple : JAPANESE UNIONS FORCE FOREIGN LABOUR BAN
(Réponse : force)

F) Rétablir dans un ordre correct quatre phrases composant normalement un texte, en étudiant le sens, les accords, les mots d'articulation, etc.

G) Répondre par VRAI ou FAUX à une série d'affirmations faites à propos d'un article de journal.

Afin de compléter cette description du test, nous avons jugé bon de réunir quelques autres renseignements de type différent sous forme d'un tableau très simplifié (tableau 1), qui complète le jugement sur un plan plus purement grammatical ou linguistique. On y voit apparaître de manière totalement *subjective* certains traits dominants de chacune des séries du test. Tout d'abord, au niveau du lexique, c'est-à-dire sémantique et morphologie. La colonne **sémantique** montre si les séries font appel à la connaissance du vocabulaire, à la compréhension du sens des mots. La colonne **morphologie** révèle si les séries obligent l'étudiant à reconnaître différentes formes de mots. Enfin, au niveau de la **syntaxe,** la 3e colonne permet de juger si l'étudiant maîtrise correctement la structure des phrases, c'est-à-dire l'arrangement des mots au sein de chaque proposition.

Pour chaque série, on trouvera deux croix (++) si l'aspect considéré est jugé primordial, une croix (+), s'il est important. Le tableau indique d'autre part si les mots, les phrases ou les documents présentés dans chaque série sont authentiques ou s'ils ont été créés pour les besoins du test.

Précisons pour terminer que le test, dans les versions 73 et 74, est noté sur 100 points, à raison de 10 points pour les exercices A, B, C, D, E, 22 points pour F et 28 points pour G, et d'autre part que l'ensemble des réponses apportées par l'étudiant tient sur une seule page, sous forme de croix pour la plupart des exercices, ce qui permet une correction extrêmement rapide à l'aide d'une grille perforée.

3.2. Présentation et exploitation du test 73

Notre exploitation a pour buts essentiels de vérifier si le test détermine réellement le profil de connaissances que nous souhaitons, et d'autre part de s'assurer que tous ses éléments sont pertinents. C'est pourquoi, outre le calcul de l'écart type et de la moyenne, nous avons jugé bon de calculer deux nouvelles valeurs : un **index de discrimination** et une **fréquence d'apparition des notes.**

3.2.1. *Index de discrimination* (IDD)

Cet index * se présente sous la forme d'un nombre compris entre — 1 et + 1. Si IDD est égal à +1, cela signifie que tous les individus ayant répondu correctement à la question envisagée sont situés parmi les meilleurs au classement général *(question discriminante)*. Si IDC est égal à 0, cela signifie qu'autant de « bons » étudiants que de « mauvais » ont répondu correctement *(question non discriminante)*. Enfin, si IDD est négatif, cela signifie que ce sont les plus faibles qui ont le mieux répondu à la question *(question aberrante)*. Le but recherché sera donc d'obtenir une valeur d'IDD positive et aussi proche de 1 que possible.

Citons à titre d'exemple : " an apple pie " (série A), IDD = 0,41 avec 384 étudiants ; l'exercice est donc relativement discriminant.

3.2.2. *Fréquence d'apparition des notes*

Ce chiffre indique pour chaque réponse, correcte ou non, le nombre d'étudiants qui l'ont choisie :

Exemple : pour " an apple pie " (série A), a : 36
b : 76
c : 248
d : 16
pas de réponse : 8 (total : 384)

L'étude simultanée de ces deux chiffres (IDD et fréquence) est fondamentale pour déterminer avec précision où se situent les difficultés des étudiants. Il est également bon de signaler que la fréquence d'apparition est totalement indépendante de l'indice de discrimination car, de deux exercices également discriminants, l'un peut être facile (grand nombre de réponses exactes), l'autre difficile (très peu de réponses exactes). Il est donc nécessaire de manier conjointement ces deux valeurs comme nous l'avons fait en 1973 après avoir fait passer le test à 120 étudiants.

* Une explication simple de cet index de discrimination a été donnée par Groot (1972). Voici la formule de calcul qu'il préconise : Si l'on divise l'ensemble des étudiants *classés* en trois groupes de même nombre,

$$IDD = \frac{\left(\begin{array}{c}\text{nombre de réponses correctes}\\ \text{dans le groupe de tête}\end{array}\right) - \left(\begin{array}{c}\text{nombre de réponses correctes}\\ \text{dans le groupe de queue}\end{array}\right)}{\text{1/3 du nombre des étudiants}}$$

3.2.3. *Amélioration du test 73*

Afin de créer la version 74 du test, nous avons systématiquement modifié ou supprimé tous les exercices non discriminants (IDD inférieur à 0,15). Lors de l'exploitation du test 74 (400 étudiants), nous avons pu vérifier que les exercices fortement discriminants en 1973 l'étaient toujours et que les exercices modifiés donnaient en général de meilleurs résultats qu'auparavant. Pour la troisième version du test (test 75), nous n'avons plus l'intention de procéder à des modifications, mais seulement à une légère miniaturisation en ne tenant compte que des exercices les plus discriminants qui apparaîtront seuls ou volontairement « noyés » dans la masse des autres.

Parallèlement à cet indice de discrimination de base (calculé pour chaque exercice), on peut également calculer un indice pour chaque série d'exercices. Le tableau 2 montre que dans la version 74 du test toutes les séries sont beaucoup plus discriminantes que dans la version 73. On note par ailleurs un maximum pour F et G qui peut s'expliquer de deux façons : ces séries ont, soit un pouvoir de discrimination effectivement supérieur à celui des autres, soit une importance relative trop élevée à l'intérieur du test (Par exemple, si le barème leur accorde trop de points, elles seront nécessairement discriminantes vis-à-vis du résultat final). La suite de l'étude permettra de choisir la conclusion la plus probable.

3.3. Exploitation et amélioration du test 74

Donc, tout en notant le comportement particulier des séries F et G, nous avons décidé d'aborder l'étude du test 74 d'une manière plus complète en introduisant le calcul d'une nouvelle valeur statistique.

3.3.1. *Coefficient de corrélation*

La statistique nous montre qu'à partir de distributions normales de notes[7], on peut calculer un **coefficient de corrélation.** Etant donné le caractère assez complexe des formules de calcul de cette valeur, nous engageons le lecteur qui désirerait de plus amples informations à ce sujet, à consulter un ouvrage de statistiques[8]. Mais nous dirons tout de même que ce coefficient peut varier entre

[7] En statistique, on entend par *population normale*, ou *gaussienne*, une population distribuée de façon symétrique par rapport à la valeur moyenne de la population, et présentant un maximum d'*individus* pour cette valeur moyenne et des minima pour les valeurs extrêmes (cf. figure 3 c, pour la répartition : *ébauche 75*).

[8] Signalons à ce propos un excellent ouvrage de Mialaret et Pham (1967) destiné aux lecteurs qui n'auraient pas une formation mathématique très avancée.

0 et + 1 selon que les répartitions sont indépendantes ou corrèlent fortement. En d'autres termes, plus la répartition ou le classement des éudiants à une épreuve est semblable à leur répartition à une autre épreuve, plus nous dirons que les épreuves corrèlent, et plus le coefficient sera proche de 1. Au contraire, dans le cas de disparité totale entre les épreuves, le coefficient sera proche de 0 et il n'y aura pas corrélation.

Nous avons donc calculé à l'aide de l'ordinateur les coefficients de corrélation, d'une part entre chaque couple de séries d'exercices, et d'autre part entre chaque série et la note finale du test (tableaux 3 et 4). Nous sommes parfaitement conscients de la prudence avec laquelle il faut manier ces chiffres du fait que la distribution de chaque série n'est pas toujours normale, en dépit de la loi des grands nombres. Toutefois, l'analogie entre les résultats obtenus en 1973 et en 1974 nous permet de tirer quelques conclusions à partir des variations relatives, et nous montrons ici comment nous pensons modifier le test en conséquence pour 1975.

Ce qui est dit dans les lignes suivantes s'applique aussi bien au tableau 3 qu'au tableau 4, c'est-à-dire à la version 73 qu'à la version 74 du test, mais il est évident que plus d'attention doit être portée à la seconde, qui d'une part est plus récente, et d'autre part a été présentée à un plus grand nombre d'étudiants (environ 400).

Ces tableaux permettent de noter que dans tous les cas, la corrélation est grande entre la note partielle d'une série d'exercices et la note globale du test. Cela signifie que chaque série est pertinente, vis-à-vis de l'aptitude à la compréhension écrite. On remarque que la série G corrèle plus fortement que les autres avec la note finale. L'explication est simple : Cette série peut apporter à elle seule jusqu'à 28 points sur 100, donc près du tiers du total. Cette proportion, qui peut paraître injustifiée ou exagérée, a été choisie délibérément du fait que G est la série se rapprochant le plus de la situation réelle de compréhension écrite (lecture d'un texte).

En ce qui concerne les séries d'exercices prises deux à deux, on note que la corrélation est plus faible mais reste moyenne, excepté pour la série F. Cela peut signifier que ces séries ne testent pas strictement les mêmes aptitudes (sans quoi les coefficients seraient plus proches de 1), mais qu'elles ne sont pas non plus totalement indépendantes (car dans ce cas, les coefficients seraient beaucoup plus faibles).

La série F, nous l'avons déjà souligné, se distingue des autres par un comportement totalement différent : elle ne corrèle que très faiblement avec chaque autre série, et par contre elle corrèle très bien avec la note finale. Deux explications sont possibles, qui rejoignent celles données à propos de l'indice de discrimination :

— F évalue peut-être un aspect de la compréhension écrite entièrement distinct de ceux apparaissant dans les autres séries ;

— ou bien au contraire, F ne teste pratiquement pas la compréhension écrite et la valeur élevée du coefficient de corrélation vis-à-vis de la note finale provient d'un mauvais barème qui lui attribue trop de points. Nous penchons a priori pour cette seconde explication, tout d'abord parce qu'il n'existe probablement pas parmi les facteurs de compréhension écrite un facteur totalement indépendant des autres : il y a toujours nécessairement interaction, recouvrement partiel entre les différents effets. D'autre part, cette série est notée sur 22 points alors que la plupart des autres séries sont notées sur 10. La série F consistant à remettre dans l'ordre des phrases préalablement mélangées, il est possible que le travail demandé fasse appel à un autre facteur que la compréhension écrite : les connaissances orales étant exclues à coup sûr, il reste le facteur expression écrite qui, effectivement joue peut-être un rôle beaucoup plus important que nous ne l'avions imaginé a priori.

Afin d'expliquer cette anomalie et de mieux juger de l'importance propre de chaque série, nous avons calculé à nouveau le coefficient de corrélation mais en omettant tour à tour une série dans le calcul de la note finale. Ces résultats apparaissent dans le tableau 5. Ainsi, on trouve dans la colonne " —A " les coefficients entre chaque série et la note finale quand on ne tient pas compte des résultats obtenus à A.

Si l'on considère d'abord les valeurs qui n'ont aucun rapport avec la série F, plusieurs conclusions s'imposent . On note que sur chaque ligne horizontale, la valeur du coefficient de corrélation ne bouge pratiquement pas. Dans presque tous les cas on a une variation négative inférieure à 4 %. Cette diminution insignifiante de la corrélation montre que la série envisagée fonctionne dans le même sens que la série momentanément supprimée, quant à l'aspect de la langue qui y est testé.

En second lieu, on constate que la note de chaque série supprimée (valeurs entre parenthèses) corrèle moins qu'avant avec la note finale correspondante (diminution de 15 à 20 %). Ce résultat est tout à fait logique puisque chaque série perd à cet instant toute influence sur le total.

Enfin la série F se comporte ici encore d'une manière opposée aux autres : le retrait d'une série quelconque fait augmenter légèrement la corrélation de F avec la note finale (lignes horizontales). F teste donc apparemment des connaissances différentes. Verticalement, on voit que le retrait de F fait augmenter d'environ 15 % la corrélation de chaque série avec la note globale. L'absence de F permet donc à ces séries de jouer un rôle plus important qui sans cela se trouve masqué. Enfin, on voit que la note F ne corrèle pratiquement plus avec la

note finale quand F n'est pas pris en compte : le coefficient est égal à 0,29 alors que pour les autres séries, il était compris entre 0,46 et 0,57.

Il apparaît donc avec évidence, et ceci confirme les conclusions précédentes, que cette série F joue un rôle radicalement opposé à celui des autres et qu'elle ne teste pas les mêmes aptitudes.

3.3.2. *Modification du test 74*

Afin d'apporter une seconde série de modifications au test, nous avons travaillé dans les directions suivantes :

— Choisir et isoler les questions les plus discriminantes.

— Réduire l'importance accordée à la série F.

— Etudier les conséquences de ces modifications sur le comportement de F.

Nous avons donc procédé tout d'abord à une réduction du test à 6 questions par série en supprimant toutes les questions ayant un faible indice de discrimination. De plus, pour G, où les réponses sont du type VRAI ou FAUX, nous avons diminué la part du hasard en proposant 9 questions à l'étudiant mais en n'accordant des points qu'à partir de la quatrième réponse exacte.

Le tableau 6 indique la valeur des coefficients de corrélation pour cette version modifiée du test. On note qu'après avoir omis les questions les moins discriminantes on a bien fait augmenter, dans presque tous les cas, la corrélation entre les couples de séries (comparer avec le tableau 4). D'autre part la corrélation entre chaque série et la not finale est maintenant mieux répartie : Les valeurs sont très proches les unes des autres. Les valeurs les plus faibles ont augmenté, la valeur trop forte pour G a diminué. Toutes les séries sauf F corrèlent unanimement avec la note finale.

En ce qui concerne la série F, on voit que sans conteste elle est à rejeter de ce test de compréhension écrite : il y a confirmation du faible taux de corrélation entre elle et les autres séries ; mais de surcroît, la nouvelle répartition des points a fait chuter très nettement le coefficient entre F et la note globale (diminution d'environ 30 %), montrant bien qu'il n'y a plus maintenant qu'un très faible rapport entre la note F obtenue par un étudiant et son résultat final au test.

Nous avons donc décidé de supprimer cette série F et nous obtenons alors une nouvelle version du test (ébauche 75) ne comportant plus que 6 séries, chacune pouvant rapporter 6 points. Le tableau 7 permet de mieux suivre l'évolution des coefficients de corrélation entre chaque série et la note finale au fur et à mesure de l'évolution du test.

En étudiant de près la liste de classement des étudiants obtenue avec la version 74 et celle obtenue avec cette nouvelle version, nous constatons que d'une manière générale, les « bons » restent bons, avec bien entendu quelques déplacements, mais que c'est plutôt du côté des faibles que des variations se font sentir, montrant que la version 74 du test accordait probablement trop de points à certains d'entre eux, soit par le biais du hasard, soit par le biais de certaines questions non discriminantes qui tendaient à leur accorder des points avec trop de facilité. D'autre part la distribution des notes est beaucoup plus normale que pour la version 74 du test (figure 3).

Cette ébauche de la version 75 du test semblant donner de meilleurs résultats, nous avons effectué un nouveau calcul des indices de discrimination pour chaque série d'exercices : Le tableau 8 indique ces résultats en reprenant ceux du tableau 2.

La ligne *ébauche 75* montre combien ces indices ont augmenté depuis la première version, jusqu'à 100 % dans certains cas !

Pour F, on a bien entendu une diminution très sensible puisque la série est maintenant supprimée et n'intervient plus dans la note finale. Il ressort bien de ce tableau que les étudiants réussissant l'épreuve F ne sont pas nécessairement les mêmes que ceux réussissant le test dans son ensemble.

Pour G, on a une augmentation sensible par rapport à la version 73 mais qui se solde en fait par une diminution relative par rapport aux autres séries. Cette variation est due à la modification du barème qui accorde moins de points à G. Nous aurons toutefois à reconsidérer ce problème lors de la mise au point finale de la version 75 afin de redonner un peu plus d'importance à cette série qui reste tout de même, nous en sommes persuadés, très valable du point de vue comprehension écrite, et qui est très proche du type d'enseignement subi par les étudiants.

Cependant, nous nous en tiendrons là pour ces modifications car nous pensons avoir suffisamment montré l'intérêt présenté par un ordinateur dans l'exploitation et la mise au point d'un test en vue d'atteindre à la fois une **réduction maximale** et un **pouvoir de discrimination optimal**.

CONCLUSION

Nous sommes donc, ici à Nancy-II, en train de faire nos premiers pas vers une utilisation de l'ordinateur dans le domaine des langues vivantes. Nous avons mentionné la possibilité d'utiliser des programmes de traitement (classement, calcul de moyenne, etc.). Nous avons aussi montré comment on peut essayer de mettre au point un matériel pédagogique (par exemple un test), afin d'en faire un outil de travail sûr et scientifiquement éprouvé. Mais nous n'avons donné ici qu'une infime partie des applications que peut offrir l'ordinateur à l'enseignant d'une langue vivante. Les perspectives de recherche sont très nombreuses : il serait par exemple à la fois utile et passionnant de se pencher sur le problème déjà mentionné dans cet article, qui consiste à comparer différentes séries de notes, par exemple lors du calcul d'une moyenne de fin d'année à partir de plusieurs notes partielles et d'une note finale, afin de pouvoir, sans se tromper, déterminer si un étudiant dont la note aurait mathématiquement baissé n'aurait pas, en réalité, fait des progrès par rapport à l'ensemble du groupe d'étudiants. Il est évident qu'une telle recherche, pour être valable, doit être entreprise avec un grand nombre d'étudiants, et l'on conçoit mal qu'elle soit menée à bien sans l'utilisation d'un ordinateur.

Nous espérons avoir montré par les quelques exemples et les quelques perspectives d'avenir que nous avons rapportés ici, combien l'ordinateur peut rendre de services à l'enseignant ou au chercheur en langue vivante. C'est pourquoi il serait souhaitable pour tous les pédagogues enseignants ou chercheurs dans cette branche, que soit faite une synthèse de l'état des recherches sur les applications de l'informatique et du traitement automatique aux langues vivantes. Un minimum de connaissance dans ce domaine pourrait avantageusement éviter à bon nombre d'entre nous de perdre un temps précieux à effectuer des travaux fastidieux et facilement automatisables, ou de se lancer dans la réalisation de matériaux risquant d'être inefficaces ou même inutilisables.

ANNEXE

Série	Sémantique	Morphologie	Syntaxe	Contenu de l'exercice
A	+		++	10 mots composés
B	+	+	++	50 mots grammaticaux (adverbes, conjonctions, adjectifs)
C	++	+	+	45 mots grammaticaux (idem)
D			++	5 phrases réelles tirées d'articles de journaux
E		++	+	10 titres de journaux, non modifiés
F			++	2 textes authentiques d'environ 70 à 80 mots chacun, tous deux donnés sous forme de 4 phrases mélangées
G	+	+	+	Article de journal non modifié, 600 à 700 mots, reproduit sous son aspect typographique original

Tableau 1 : Aspects linguistiques des différentes séries d'exercices du test (version 73)

	A	B	C	D	E	F	G
IDD 73	0,54	0,36	0,36	0,36	0,38	0,46	0,59
IDD 74	0,64	0,54	0,64	0,67	0,71	0,79	0,82

Tableau 2 : Evolution de l'indice de discrimination de chaque série d'exercices en fonction des modifications apportées au test.

	B	C	D	E	F	G	TOTAL
A	0,33	0,27	0,25	0,33	0,17	0,26	0,53
B		0,32	0,10	0,25	0,08	0,08	0,41
C			0,32	0,35	0,14	0,17	0,51
D				0,48	0,09	0,33	0,50
E					0,13	0,25	0,54
F						0,21	0,70
G							0,64

Tableau 3 : Coefficients de corrélation pour le test 73

	B	C	D	E	F	G	TOTAL
A	0,39	0,39	0,43	0,46	0,09	0,40	0,56
B		0,42	0,43	0,49	0,19	0,48	0,64
C			0,36	0,45	0,34	0,42	0,66
D				0,46	0,15	0,43	0,62
E					0,20	0,45	0,65
F						0,27	0,65
G							0,78

Tableau 4 : Coefficients de corrélation pour le test 74

	TOTAL	-A	-B	-C	-D	-E	-F	-G
A	0,56	(0,46)	0.55	0,55	0,54	0,54	0,64	0,55
B	0,64	0,63	(0,54)	0,63	0,62	0,62	0,70	0,62
C	0,66	0,65	0,65	(0,57)	0,66	0,65	0,65	0,68
D	0,62	0,60	0,61	0,62	(0,49)	0,60	0,70	0,61
E	0,65	0,64	0,64	0,65	0,64	(0,56)	0,72	0,65
F	0,65	0,69	0,68	0,66	0,69	0,68	(0,29)	0,74
G	0,78	0,78	0,78	0,79	0,78	0,78	0,84	(0,56)

Tableau 5 : Test 74 : variation des coefficients de corrélation entre chaque série et la note finale, lorsqu'une série est retirée du test.
Dans la colonne —A, on ne tient pas compte de la série A pour le calcul de la note finale.

	B	C	D	E	F	G	TOTAL
A	0,43	0,44	0,44	0,54	0,13	0,43	0,71
B		0,47	0,43	0,49	0,15	0,41	0,71
C			0,39	0,47	0,25	0,46	0,72
D				0,47	0,13	0,41	0,66
E					0,16	0,49	0,75
F						0,28	0,46
G							0,72

Tableau 6 : Test 74 modifié · Coefficient de corrélation pour le test normalisé à 6 points par série.

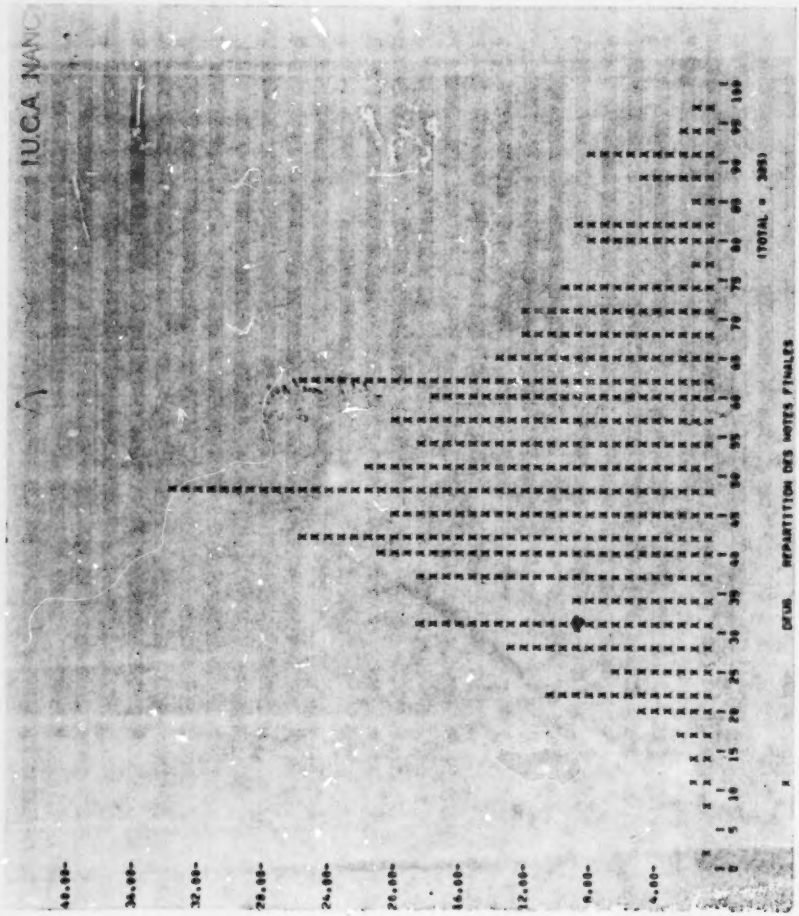
	A	B	C	D	E	F	G	Référence
Total 73	0,53	0,41	0,51	0,50	0,54	0,70	0,64	tableau 3
Total 74	0,56	0,64	0,66	0,62	0,65	0,65	0,78	tableau 4
Total 74 normalisé	0,71	0,71	0,72	0,66	0,75	0,46	0,72	tableau 6
Total ébauche 75	0,74	0,73	0,73	0,68	0,78	——	0,74	——

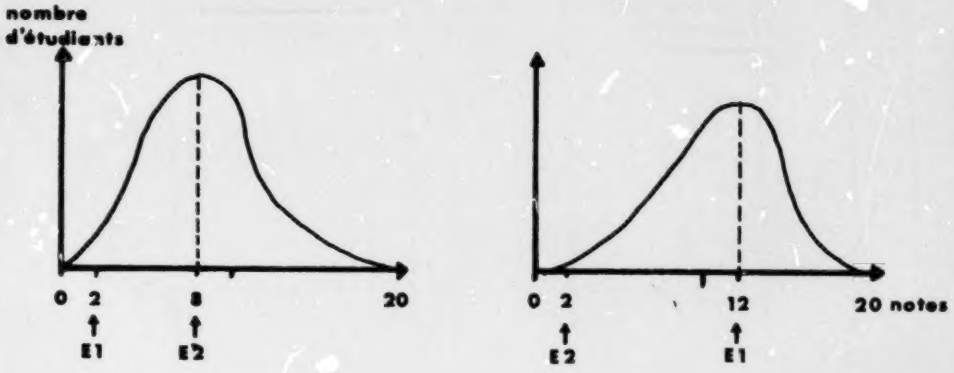
Tableau 7 : Variation des coefficients de corrélation entre la note finale et les différentes séries pour chaque nouvelle version du test.

	A	B	C	D	E	F	G
Test 73	0,54	0,36	0,36	0,36	0,38	0,46	0,59
Test 74	0,64	0,54	0,64	0,67	0,71	0,79	0,82
Ebauche 75	0,75	0,72	0,74	0,69	0,78	0,26	0,65

Tableau 8 : Evolution de l'indice de discrimination de chaque série d'exercices en fonction des modifications apportées au test.

Figure 1 : Exemple de diagramme ou d'histogramme obtenu à l'aide d'un ordinateur. En abscisse, les notes ; en ordonnée, le nombre d'étudiants.





Moyenne de l'étudiant E1 : $\frac{2 + 12}{2} = 7/20$

Moyenne de l'étudiant E2 : $\frac{8 + 2}{2} = 5/20$

Figure 2 : Exemples de répartitions différentes d'étudiants en fonction de leurs notes à deux épreuves (A et B).

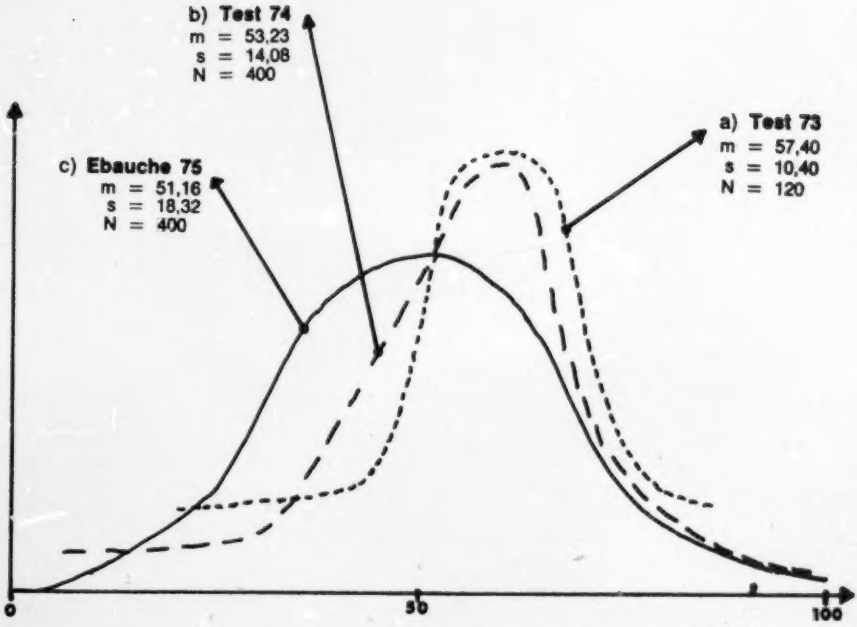


Figure 3 : Evolution de la répartition des notes du test de classement m = moyenne arithmétique, s = écart type, N = nombre d'étudiants, échelle verticale variable.

SUMMARY

The aim of this paper is to present linguists — both members of research teams and language teachers — with some possible applications of computing science to language teaching. First, it is possible to use a computer's output system merely in order to print any type of information, e.g. a list of students' names and their marks. Such a list of results may be remodelled using very simple and very numerous classifying programmes. The computer also allows the teacher to make more elaborate calculations and obtain very useful results such as the arithmetic mean, the standard deviation and so on, whatever the number of students. Such results permit a more thorough understanding and judgement of the marking scheme, and a better " visualisation " of the distribution of the students, which proves to be very useful when making comparison of the students' levels.

To illustrate clearly these applications, a concrete example is given concerning the improvement of a reading comprehension test. The test is fully described and its evaluation over three years is presented and explained : modifications of all sorts were introduced in accordance with the interpretation of mathematical and statistical data. The discrimination factor makes it clear whether any particular item discriminates effectively ; the frequency factor shows the number of answers — right or wrong — given for each exercise. Finally, the corrolation factor shows whether all the different series of exercices really contribute in testing the same kind of reading skill.

All these data are plotted or given in charts, and are thoroughly discussed in order to explain how it is possible to transform such skill-testing material to make it as concise as possible while giving it as great as a discriminatory power as possible.

REFERENCES BIBLIOGRAPHIQUES

GROOT, G. (1972). « Le problème de l'évaluation dans l'enseignement des langues vivantes » In *Cycle de linguistique appliquée à la pédagogie des langues*, Volume 3, pp. 153-154. Conférence de Hasselt, 1972.

LONCHAMP, F. (1971). « Analyse factorielle de tests de compréhension orale et écrite » In *Mélanges Pédagogiques*. C.R.A.P.E.L., Université de Nancy-II.

MIALARET, G. et PHAM, D. (1967). *Statistique à l'usage des éducateurs*. Paris, Presses Universitaires de France.